CHAPTER 8 - Limit Theorems

8.1- Inhoduchin

The most inputernt theoretical
result in probability theory are
limit theorems.

Loof these: laws of large numbers,
central limit theorem #

- Plevens are considered to be lows of large numbers if they are concerned with stating conditions under which the average of a segmence of a random variables converges for the expected average

- By contast, central limit theorems are concerned with determining.

conditions under which the sum of a

large number of undom variables has a proherbility dishibition.
That is approximately would.

8.2-CHEBYSHEV'S INFQUALITY and the WEAK LAW OF LARGE NUMBERS

- We start this section by proving a result known as Markov's regulation

Proposition 2.1. Markov's inequality

If \hat{X} is a random variable that takes only nonnegative values, then, for any variable a > 0,

$$P\{X \ge a\} \le \frac{E[X]}{a}$$

Proof. For a > 0, let

$$I = \begin{cases} 1 & \text{if } X \ge a \\ 0 & \text{otherwise} \end{cases}$$

and note that, since $X \ge 0$,

$$I \le \frac{X}{a}$$

Taking expectations of the preceding inequality yields

$$E[I] \le \frac{E[X]}{a}$$

which, because $E[I] = P\{X \ge a\}$, proves the result.

As a corollary, we obtain Proposition 2.2.

Proposition 2.2. Chebyshev's inequality

If X is a random variable with finite mean μ and variance σ^2 , then, for any value k > 0,

$$P\{|X - \mu| \ge k\} \le \frac{\sigma^2}{k^2}$$

Proof. Since $(X - \mu)^2$ is a nonnegative random variable, we can apply Markov's inequality (with $a = k^2$) to obtain

$$P\{(X - \mu)^2 \ge k^2\} \le \frac{E[(X - \mu)^2]}{k^2}$$
 (2.1)

But since $(X - \mu)^2 \ge k^2$ if and only if $|X - \mu| \ge k$, Equation (2.1) is equivalent to

$$P\{|X - \mu| \ge k\} \le \frac{E[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

and the proof is complete.

Hohe importance of Markov's and Chebysher's inequalities is that they enable us to derive bounds on probabilities when only the near, or both the vean and the variance, of the probability dishibution are known.

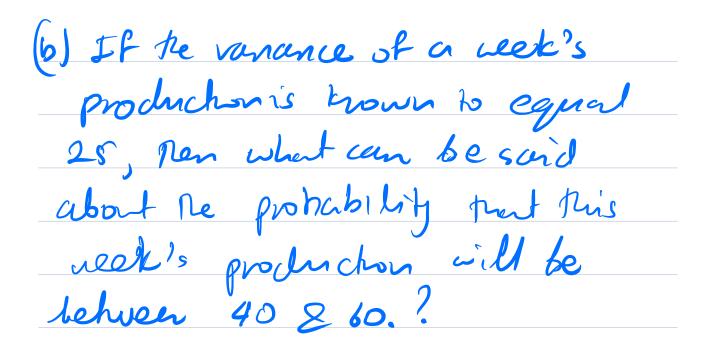
EXAMPLE 2a

Suppose that it is known that the number of items produced mid hetry during a reek is a random variable with near 50.

a) What can be said about the probability that this neek's production will exceed 75?

-P det X be te number of itens that will be produced in a week

Markov's inequality:



-P by (he by sher's regnerlitj.

PUX-50 1>10) \(\frac{\sigma^2}{10^2} = \frac{1}{4}

Ikno:

P(1x-10) 2 1-1=3

So the probability part this week's production will be better 40 & 60 is at least 75%.

- As Chebysher's requality is valid for all dishibutions of the vanction variable. X, we cannot expect the bound on the probability to be very close to the actual probability in most cases

Example 26

If $\chi \sim U(0, 0)$, then $E[\chi]_{=5}$ & $Var(\chi) = \frac{2r}{3}$

it hollars hom Che bysher's requality

 $P(1x-51>4) \leq 25 = 0.52$ 306)

whereas he exact result is

P(1x-51>4)=6.20

Although Chebysher? requality is correct the upper bound that it provides is not particularly close to the actual probability.

Similarly, if X is a normal random variable with near for & varance or Chehysher's inequality states that

 $PCIX-\muI>2\sigma)\leq \frac{1}{4}$

Me achial probability is gien

$$P\{|X - \mu| > 2\sigma\} = P\left\{ \left| \frac{X - \mu}{\sigma} \right| > 2 \right\} = 2[1 - \Phi(2)] \approx .0456$$

Note:

Che by sher's requality is often used as a newchical book of proving result. Lo This is Illustrated. Birst by Roposition 2.3 and pen, nost uportally by he weak—law of large numbers

Proposition 2.3. If Var(X) = 0, then

$$P\{X = E[X]\} = 1$$

In other words, the only random variables having variances equal to 0 are those which are constant with probability 1.

Theorem 2.1 The weak law of large numbers

Let $X_1, X_2, ...$ be a sequence of independent and identically distributed random variables, each having finite mean $E[X_i] = \mu$. Then, for any $\varepsilon > 0$,

$$P\left\{ \left| \frac{X_1 + \dots + X_n}{n} - \mu \right| \ge \varepsilon \right\} \to 0 \quad as \quad n \to \infty$$

Proof. We shall prove the theorem only under the additional assumption that the random variables have a finite variance σ^2 . Now, since

$$E\left[\frac{X_1 + \dots + X_n}{n}\right] = \mu$$
 and $Var\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{\sigma^2}{n}$

it follows from Chebyshev's inequality that

$$P\left\{\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \ge \varepsilon\right\} \le \frac{\sigma^2}{n\varepsilon^2}$$

and the result is proven.

8.3- Re Central Limit Theven



-In simple words:

Lo It skiles that the sum of a large number of independent random vanables has a distribution that is approximetly normal. Hence, if not only provides a suple nethod by computing approximate probabilités Br sums of indépendent random vanables, but also helps explain the remarkable fact that Le empirical Reguencies of so many natural populations exhibit bell-shaped (normal) cures.

Theorem 3.1 The central limit theorem

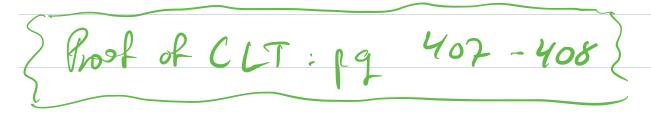
Let X_1, X_2, \ldots be a sequence of independent and identically distributed random variables, each having mean μ and variance σ^2 . Then the distribution of

$$\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal as $n \rightarrow \infty$. That is, for $-\infty < a < \infty$,

$$P\left\{\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \le a\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-x^2/2} \, dx \quad as \quad n \to \infty$$

The key to the proof of the central limit theorem is the following lemma, which we state without proof.



EXAMPLE 3a

An astronomer is interested in measuring the distance, in light-years, from his observatory to a distant star. Although the astronomer has a measuring technique, he knows that, because of changing atmospheric conditions and normal error, each time a measurement is made it will not yield the exact distance, but merely an estimate. As a result, the astronomer plans to make a series of measurements and then use the average value of these measurements as his estimated value of the actual distance. If the astronomer believes that the values of the measurements are independent and identically distributed random variables having a common mean d (the actual distance) and a common variance of 4 (light-years), how many measurements need he make to be reasonably sure that his estimated distance is accurate to within $\pm .5$ light-year?

See 19 409-411) Sohn

It U1, ..., In are Pen-neconnent -D

$$Z_n = \frac{\sum_{i=1} X_i - nd}{2\sqrt{n}}$$

8.4- The Shong Law of Large Number

Lo Chis is probably pe best known result in probability treong.

-D It states that the average of a segmence of independent random variables having a common distribution will, with probability 1, converge to the mean of that distribution.

Theorem 4.1 The strong law of large numbers

Let $X_1, X_2, ...$ be a sequence of independent and identically distributed random variables, each having a finite mean $\mu = E[X_i]$. Then, with probability 1,

$$\frac{X_1 + X_2 + \dots + X_n}{n} \to \mu \qquad as \qquad n \to \infty^{\dagger}$$

As an application of the strong law of large numbers, suppose that a sequence of independent trials of some experiment is performed. Let E be a fixed event of the experiment, and denote by P(E) the probability that E occurs on any particular trial. Letting

 $X_i = \begin{cases} 1 & \text{if } E \text{ occurs on the } i \text{th trial} \\ 0 & \text{if } E \text{ does not occur on the } i \text{th trial} \end{cases}$

we have, by the strong law of large numbers, that with probability 1,

$$\frac{X_1 + \dots + X_n}{n} \to E[X] = P(E) \tag{4.1}$$

Since $X_1 + \cdots + X_n$ represents the number of times that the event E occurs in the first n trials, we may interpret Equation (4.1) as stating that, with probability 1, the limiting proportion of time that the event E occurs is just P(E).

Although the theorem can be proven without this assumption, our proof of the strong law of large numbers will assume that the random variables X_i have a finite fourth moment. That is, we will suppose that $E[X_i^4] = K < \infty$.

Proet can be bound pg 415)

8-5-0THER INEQUALITIES

20 We are sonehues conformted with schahris in which we are interested in obtuning an apper bound by a prohability of the kirm PCX-pza), where a is some positive value and when only the nean $\mu = E(x) & varance$ o22 Ver (20) of the dishrbution of X are known.

- Of course, since x-42a>0 ruphes that 12-pl > a

Co Che bysher's megnality

$$P\{X - \mu \ge a\} \le P\{|X - \mu| \ge a\} \le \frac{\sigma^2}{a^2}$$
 when $a > 0$

-D Check ort Proposition 5-1 pg 418 Proposition 5-2 pg 422 Proposition 5-3 pg 424

SUMMARY

SUMMARY

Two useful probability bounds are provided by the *Markov* and *Chebyshev* inequalities. The Markov inequality is concerned with nonnegative random variables and says that, for *X* of that type,

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

for every positive value a. The Chebyshev inequality, which is a simple consequence of the Markov inequality, states that if X has mean μ and variance σ^2 , then, for every positive k,

$$P\{|X - \mu| \ge k\sigma\} \le \frac{1}{k^2}$$

The two most important theoretical results in probability are the *central limit theorem* and the *strong law of large numbers*. Both are concerned with a sequence of independent and identically distributed random variables. The central limit theorem says that if the random variables have a finite mean μ and a finite variance σ^2 , then the distribution of the sum of the first n of them is, for large n, approximately that of a normal random variable with mean $n\mu$ and variance $n\sigma^2$. That is, if X_i , $i \ge 1$, is the sequence, then the central limit theorem states that, for every real number a,

$$\lim_{n \to \infty} P\left\{ \frac{X_1 + \dots + X_n - n\mu}{\sigma \sqrt{n}} \le a \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-x^2/2} dx$$

The strong law of large numbers requires only that the random variables in the sequence have a finite mean μ . It states that, with probability 1, the average of the first n of them will converge to μ as n goes to infinity. This implies that if A is any specified event of an experiment for which independent replications are performed, then the limiting proportion of experiments whose outcomes are in A will, with probability 1, equal P(A). Therefore, if we accept the interpretation that "with probability 1" means "with certainty," we obtain the theoretical justification for the long-run relative frequency interpretation of probabilities.